W. Dür, G. Vidal and J.I. Cirac
Institut für Theoretische Physik, Universität Innsbruck,A-6020 Innsbruck, Austria
(February 1, 2001)

We analyze the problem of quantum data compression of commuting density operators in the visible case. We show that the lower bound for the compression factor given by the Levitin–Holevo function is reached by providing an explicit protocol.

03.67.-a, 03.65.Bz, 03.65.Ca, 03.67.Hk

I. INTRODUCTION

The applications of Quantum Mechanics in the fields of communication, computation, and precision measurements are based on the possibility of encoding and manipulating information using quantum states. Thus, one of the most relevant questions in this context is the extension of Shannon's noiseless coding theorem [1] to the quantum domain. That is, to find out the minimum amount of resources needed for a faithful storage (encoding) and retrieval (decoding) of quantum states, or, equivalently, the most economical way of compressing them. For pure states, this problem was stated and solved by Schumacher [2–4]. For mixed states, however, this is still an open problem [5–7].

The problem of quantum data compression can be formulated as follows. Alice has a (stationary memoryless) quantum source that produces systems in the state (described by the density operator) ρ_k with probability p_k, where $k = 1, 2, \ldots, L$ (L finite). Let us consider a sequence $\mathcal{K}$ of N systems which, for simplicity, we consider to be qubits, created by the source. Let us denote by $\sigma^A_{\mathcal{K}} \equiv \rho_{k_1} \otimes \rho_{k_2} \ldots \otimes \rho_{k_N}$ the corresponding state [8]. Alice wants to transmit such a state to Bob by using as few qubits as possible. That is: (i) she *encodes* the sequence in a set of M qubits (i.e., with the help of her sequence she prepares them in some state) and sends them to Bob; (ii) he *decodes* the state (i.e. with the help of the qubits he has received he prepares a sequence of N systems in some state $\sigma^B_{\mathcal{K}}$). The goal is to find the procedure for which, for sufficiently long sequences, Bob's state $\sigma^B_{\mathcal{K}}$ is "arbitrarily close" to $\sigma^A_{\mathcal{K}}$ and, at the same time, M is minimal (arbitrarily close means with respect to some measure of fidelity, see below). The quantity $C = \lim_{N\to\infty} M/N$ is called compression factor.

In the case where the ρ_k correspond to pure states one finds that $C = S(\rho)$ [2], where

$$\rho \equiv \sum_{k=1}^{L} p_k \rho_k, \tag{1}$$

and

$$S(\rho) \equiv -\mathrm{tr}[\rho \log_2(\rho)], \tag{2}$$

is the von Neumann entropy of ρ. When the ρ_k correspond to mixed states, however, the value of C is not known (except for the somehow simple case in which the supports of the operators ρ_k are orthogonal [9]). It can be shown that $S(\rho) \geq C \geq I(\{p_k\}, \{\rho_k\})$ [5,6], where

$$I(\{p_k\}, \{\rho_k\}) = S(\rho) - \sum_k p_k S(\rho_k), \tag{3}$$

is the Levitin–Holevo function. In Ref. [6], the authors analyze several cases where they are able to show that $S(\rho) > C$ by providing explicit protocols. However, none of those protocols achieve the lower bound $I(\{p_k\}, \{\rho_k\})$. Thus, the question whether this limit can be reached or not is still open. In fact, it has been argued [7] that in the affirmative case one could assign a definite meaning to the Levitin–Holevo function besides the well known one related to the maximum amount of classical information that can be stored and retrieved in and from quantum states [10–12].

There are two different scenarios where quantum data compression of mixed states has been analyzed [5–7]. In the so–called *visible* scenario, Alice knows the state $\sigma^A_{\mathcal{K}}$ she wants to compress. In the *blind* one, she does not know it. Obviously, the compression factor in the visible scenario is smaller than or equal to that in the latter one. In particular, for pure states both compression factors coincide [2].

In this paper we study the compression of quantum mixed states in the visible scenario, and in the case in which the operators ρ_k commute with each other. We provide an explicit protocol which reaches the lower bound for the compression factor, which implies that

$$C = I(\{p_k\}, \{\rho_k\}). \tag{4}$$

The basic idea to achieve such compression factor is to let Alice and Bob change the encoding/decoding procedure randomly from sequence to sequence. For that, we will assume that Alice and Bob possess the same random number generator (or, equivalently, that they share a list of random numbers). We will concentrate on the case in which the systems under consideration are qubits. As we will indicate, the generalization to higher dimensional systems is straightforward. Note that, as shown in Ref. [6], the problem analyzed in this paper is equivalent to the one of classical data compression of probability distributions. We will nevertheless use a quantum mechanical

language in view of a possible extension of our protocol to the case in which the operators ρ_k do not commute. On the other hand, the reason why our protocol achieves the compression factor (4) can be easily understood in terms of typical subspaces (or typical sequences in the classical case). Thus, we will first explain how our protocol works by using this concept. Once this is clear, a detailed proof can be easily constructed. It has come to our attention that [13] presents an alternative proof of the achievability of (4) using rate distortion theory.

This paper is organized as follows. In Section II we qualitatively explain our protocol using the concept of typical sequences. In Section III we describe in detail our protocol for the case of two states ($L = 2$) and show that it achieves the compression factor (4). The protocol can be straightforwardly generalized to $L > 2$ by following the ideas of Section II. However, we do not include the detailed proofs here since they require an involved notation, and do not add any new idea to the problem. In Section IV we discuss possible extensions of our protocol. Finally, the Appendix is concerned with some technical details.

II. DESCRIPTION IN TERMS OF TYPICAL SEQUENCES

In this section we formulate the problem in terms of typical sequences, which allows us to explain the basic idea of our protocol. We assume that Alice wants to send a sequence of N qubits to Bob, each one in state ρ_k with probability p_k, where all the ρ_k commute. We can always write

$$\rho_k = \lambda_k |1\rangle\langle 1| + (1-\lambda_k)|0\rangle\langle 0|, \tag{5}$$

Thus, we have

$$\rho = \sum_{k=1}^{L} p_k \rho_k = \overline{P}_1 |1\rangle\langle 1| + \overline{P}_0 |0\rangle\langle 0|, \tag{6}$$

where

$$\overline{P}_1 = \sum_{k=1}^{L} p_k \lambda_k, \quad \overline{P}_0 = 1 - \overline{P}_1. \tag{7}$$

These quantities are the probability that the quantum source creates the state $|1\rangle$ and $|0\rangle$, respectively.

As mentioned in the introduction, the goal is to compress a sequence of the form $\sigma_{\mathcal{K}}^A \equiv \rho_{k_1} \otimes \rho_{k_2} \ldots \otimes \rho_{k_N}$, where $k_i = 1, 2, \ldots, L$. We will denote by v_k a vector whose elements indicate the positions at which the operator ρ_k appears. For example, if we take the sequence

$$\underbrace{\rho_1 \otimes \rho_1 \ldots \rho_1}_{n_1} \otimes \underbrace{\rho_2 \otimes \rho_2 \ldots \rho_2}_{n_2} \otimes \ldots \otimes \underbrace{\rho_L \otimes \rho_L \ldots \rho_L}_{n_L} \tag{8}$$

then $v_1 = (1, 2, \ldots, n_1)$, $v_2 = (n_1+1, n_1+2, \ldots, n_1+n_2)$, etc.

If the sequence is sufficiently long, $\sigma_{\mathcal{K}}^A$ will contain the state ρ_k approximately $\overline{n}_k \equiv N p_k \gg 1$ times. Let us call a sequence which exactly contains such a number of times these operators "typical sequence". Moreover, since $\overline{n}_k \gg 1$ we can also apply the same idea within the sequence that Alice wants to send. If we write the operator $\sigma_{\mathcal{K}}^A$ in the basis $\{|i_1\rangle \otimes |i_2\rangle \ldots |i_N\rangle\}$ ($i_j = 0, 1$), most of the contribution will come from states with approximately $\overline{n}_k \lambda_k$ ones (and $\overline{n}_k(1-\lambda_k)$ zeros) at the positions v_k. Let us call "typical states" those with exactly such numbers of zeros and ones at the positions specified by v_k. Thus, let us concentrate on a method in which, given a typical sequence, Alice sends Bob enough information so that he can create at random one of the corresponding typical states. It is intuitively clear that if Alice can accomplish this task with $M \sim NI(\{p_k\}, \{\rho_k\})$ qubits, then she will also be able to send most of the sequences with this amount of qubits and high fidelity.

So, let us now assume that Alice and Bob use their random number generator to create *the same* random state of N qubits, each of them in the state $|0\rangle$ or $|1\rangle$ according to the probabilities $\overline{P}_0$ and $\overline{P}_1$, respectively. Let us denote by p the probability that such a state is a typical one for a given typical sequence. In that case, if they create (instead of one) $\sim 1/p$ such random states, the probability that among them there is a typical one will be very close to one. In that case, Alice just has to tell Bob which of those states randomly generated is the one that corresponds to the typical sequence she is intending to send. The number of qubits to give that information to Bob is $M = \log_2(1/p)$. Since

$$p = \frac{\left(\begin{array}{c} \overline{n}_1 \\ \overline{n}_1\lambda_1 \end{array}\right)\left(\begin{array}{c} \overline{n}_2 \\ \overline{n}_2\lambda_2 \end{array}\right)\cdots\left(\begin{array}{c} \overline{n}_L \\ \overline{n}_L\lambda_L \end{array}\right)}{\left(\begin{array}{c} N \\ N\overline{P}_1 \end{array}\right)} \tag{9}$$

we obtain that $M = \log_2(1/p) \sim NI(\{p_k\}, \{\rho_k\})$ (for $N \gg 1$).

III. PROTOCOL FOR TWO STATES

In this Section we give the protocol to achieve the compression factor (4). We will show that for any $\epsilon, \delta > 0$ there exists an N_0 such that the sequences with $N > N_0$ qubits can be encoded in $N[I(p_k, \rho_k) + \delta]$ qubits with a fidelity $\overline{F} > 1 - \epsilon$. Here, $\overline{F}$ is the averaged fidelity

$$\overline{F} = \sum_{\mathcal{K}} P_{\mathcal{K}} F(\sigma_{\mathcal{K}}^A, \sigma_{\mathcal{K}}^B), \tag{10}$$

$P_{\mathcal{K}}$ is the probability that Alice sends the sequence $\mathcal{K}$, and [14]

$$F(\sigma_1, \sigma_2) \equiv \mathrm{tr}\left[\sigma_1^{1/2}\sigma_2\sigma_1^{1/2}\right]^{1/2} = \mathrm{tr}[\sigma_1^{1/2}\sigma_2^{1/2}], \tag{11}$$

where the last equality holds for commuting operators.

We will concentrate in the case where there are only two possible states ($L = 2$). The general case can be analyzed in the same way as here, although the notation becomes much more involved. Thus, let us assume that Alice wants to send the sequence $\mathcal{K}$, consisting of N qubits in states ρ_1 or ρ_2, to Bob. As before, we will call $n_{1,2}$ (where $n_2 = N - n_1$) the number of times the operator $\rho_{1,2}$ appears in the sequence, and $v_{1,2}$ the positions where it appears. Note that in all these quantities we should write a subscript $\mathcal{K}$ indicating their dependence on the particular sequence Alice is trying to send. In order to keep the notation simple, and whenever it is clear from the context, we will omit in all the quantities the dependence on the particular sequence. On the other hand, in the protocol given below we will consider that Alice sends classical bits to Bob. Obviously, these classical bits can in turn be encoded in the same number of qubits if we choose the states $|0\rangle$ and $|1\rangle$. The protocol consists of the following encoding and decoding procedures:

1. Encoding:

   (a) Alice selects two integer numbers x_1 and x_2, with $0 \leq x_i \leq n_i$ according to the following binomial distributions

   $$P(x_i) = \left(\begin{array}{c} n_i \\ x_i \end{array} \right) \lambda_i^{x_i} (1-\lambda_i)^{n_i - x_i}, \qquad (12)$$

   where $i = 1, 2$.

   (b) Using the common random number generator Alice creates S random sequences of N bits each. Each of the bits is set to 1 or 0 according to the probability $P_1 = (\lambda_1 n_1 + \lambda_2 n_2)/N$, $P_0 = 1 - P_1$, respectively. She associates a number between 1 and S with each sequence.

   (c) If among the S sequences there are one or more with exactly x_i ones and $n_i - x_i$ zeros at the positions indicated by v_i for both $i = 1, 2$, then she chooses one of them randomly and sends the number associated with that sequence to Bob. Otherwise, she sends the number 0 (which indicates an error). Note that for that she uses $[\log_2(S+1) + 1]$ bits, where $[\ldots]$ denotes the integer part.

   (d) She also encodes in a set of $[\log_2 N + 1]$ bits the value of n_1 and sends it to Bob.

2. Decoding

   (a) Bob uses the random number generator to create the same S random sequences as Alice and assigns the same numbers. Note that Bob knows the values n_1 (since it has been sent by Alice) and $n_2 = N - n_1$.

   (b) Using the bits sent by Alice, he identifies the random sequence and prepares N qubits in the corresponding state (i.e. prepares the qubits in states $|0\rangle$ or $|1\rangle$ if the sequence contains a zero or a one at each position). If he receives the error state, he prepares the qubits in a fixed state $\sigma_0 = \mathbb{1}/2^N$.

Before showing that the above protocol achieves the desired bound, let us make some remarks. Firstly, we can replace the condition imposed by $\delta > 0$ on the number of bits needed to encode the sequences by requiring that

$$\log_2(S) = N[I(p_k, \rho_k) + f_N], \qquad (13)$$

where $f_N \to 0$ as $N \to \infty$. Note that the number of bits needed to transmit the value of n_1 can be included in f_N since $\log_2(N+1)/N \to 0$, and therefore need not be considered. Actually, one can devise a similar encoding and decoding scheme in which this number need not be transmitted. However, our scheme allows for a simpler proof of our statements. Secondly, as it is shown in the Appendix, we can replace the condition imposed by ϵ on the averaged fidelity by

$$E \equiv \sum_{n_1=0}^{N} P_{n_1} E_{n_1} < \epsilon \qquad (14)$$

where P_{n_1} is the probability that we have a sequence with exactly n_1 times ρ_1 and the rest ρ_2, and E_{n_1} is the probability that Alice sends the error bit 0 if she had one of such sequences. Thirdly, we will deal with several binomial distributions, which have the form

$$Q_y \equiv \left(\begin{array}{c} n \\ y \end{array} \right) p^y (1-p)^{n-y}, \qquad (15)$$

where $0 < p < 1$. We will use the following properties of such distribution: (i) for all $\epsilon > 0$ and $0 < \eta < 1/2$, there exists some n_0 such that if $n > n_0$ then

$$\sum_{y=[pn-n^{1/2+\eta}]}^{[pn+n^{1/2+\eta}]} Q_y > 1 - \epsilon. \qquad (16)$$

This property allows us to restrict the allowed values of the parameters. For the sake of definiteness we will take $\eta = 0.1$. (ii) For n sufficiently large and $y \in [pn - n^{1/2+\eta}, pn + n^{1/2+\eta}]$

$$Q_y > \frac{1}{2} \frac{e^{-(y-np)^2/[2(np(1-p)]}}{\sqrt{2\pi np(1-p)}}. \qquad (17)$$

Now, let us show that the protocol given above fulfills the desired properties. First, given the fact that P_{n_1} follows a binomial distribution, we can restrict the summation in (14) to the values

$$n_1 \in [\overline{n}_1 - N^{1/2+0.1}, \overline{n}_1 + N^{1/2+0.1}]. \qquad (18)$$

Moreover, the remaining sum is smaller than the maximum value of E_{n_1} where n_1 lies in the interval indicated

in Eq. (18). This value can be determined with the help of Eq. (A6). Since $P(x_{1,2})$, the probability that Alice selects the values x_1 and x_2 in the step 1(a), is a product of two binomial distributions, again for sufficiently large N we can restrict the sums to

$$x_i \in [n_i\lambda_i - N^{1/2+0.1}, n_i\lambda_i + N^{1/2+0.1}], \quad i = 1, 2. \quad (19)$$

Thus, the problem is reduced to showing that for any $\epsilon > 0$, for sufficiently large N we can choose S fulfilling (13) and so that the maximum value of $E(x_{1,2}, v_{1,2})$ with the restrictions (18) and (19) is smaller than ϵ, and where $E(x_{1,2}, v_{1,2})$ is the probability that the error state is produced given the values of $x_{1,2}$ and $v_{1,2}$ (see Appendix). We can always write $E(x_{1,2}, v_{1,2}) = [1 - R(x_{1,2}, v_{1,2})]^S$, where $R(x_{1,2}, v_{1,2})$ is the probability that if we take a sequence of zeros and ones according to the probabilities $P_{1,0}$, the sequence exactly contains x_i ones (and the rest zeros) at positions v_i, for both $i = 1, 2$. Such a probability can be calculated as $R(x_{1,2}, v_{1,2}) = Q(x_1 + x_2)P(x_{1,2}, v_{1,2}/x_1 + x_2)$, where $Q(x_1 + x_2)$ is the probability that the sequence contains $x_1 + x_2$ ones and $P(x_{1,2}, v_{1,2}/x_1 + x_2)$ is the probability that those are at the correct positions. The first one is given again by a binomial distribution; by using Eq. (17) one can easily find [16] that

$$Q(x_1 + x_2) \geq K\frac{e^{-\alpha N^{0.2}}}{\sqrt{N}} \equiv \frac{1}{a_N}, \quad (20)$$

where K and α are constants (independent of N). On the other hand

$$P(x_{1,2}, v_{1,2}/x_1 + x_2) = \frac{\left(\begin{array}{c} n_1 \\ n_1\lambda_1 \end{array}\right)\left(\begin{array}{c} n_2 \\ n_2\lambda_2 \end{array}\right)}{\left(\begin{array}{c} N \\ NP_1 \end{array}\right)} \geq 2^{-NI(\{p_k\},\{\rho_k\})-N^{1/2+0.2}} \equiv \frac{1}{b_N} \quad (21)$$

for sufficiently large N, as can be checked using the bounds given by Stirling formulas. By choosing $S = Na_Nb_N$ we obtain that $E(x_{1,2}, v_{1,2}) = [1 - R(x_{1,2}, v_{1,2})]^S \leq [1 - 1/(a_Nb_N)]^{Na_Nb_N} \to 0$ and (13) with $f_N = [\log_2(N) + o(N^{0.7})]/N \to 0$ for $N \to \infty$, as required.

IV. POSSIBLE EXTENSIONS

A. d-level systems

One can easily generalize our results to d–level systems. In that case, a quantum source produces d–level systems (qudits) in the state (described by the density operator) ρ_k with probability p_k. For a faithful transmission of N of those systems, M qudits (equivalently $M\log_2(d)$ qubits) are required. In case all ρ_k commute, the compression factor $C = \lim_{N\to\infty} M/N$ turns out to be $C = I(\{p_k\}, \{\rho_k\})/\log_2(d)$, where $I(\{p_k\}, \{\rho_k\})$ is given in (3) and the factor $\log_2(d)$ appears because we are dealing with d–level systems now. The number of qubits per signal states required for a faithful transmission is thus again given by the Levitin-Holevo function $I(\{p_k\}, \{\rho_k\})$, so the lower bound can be reached also when dealing with d–level systems.

This can be understood qualitatevely in a similar way as in the qubit case (see Sec. II). The condition that all ρ_k commute implies that we can always write

$$\rho_k = \sum_{j=1}^{d} \lambda_j^k |j\rangle\langle j|. \quad (22)$$

and thus

$$\rho = \sum_{k=1}^{L} p_k\rho_k = \sum_{j=1}^{d} \overline{P_j}|j\rangle\langle j|, \quad (23)$$

where $\overline{P_j} = \sum_{k=1}^{L} p_k\lambda_j^k$. Proceeding in the same vain as in the qubit case, we find that the "typical states" of a certain (typical) sequence have exactly $Np_k\lambda_j^k$ states $|j\rangle$ at the positions v_k. It is straightforward to calculate the probability p that a state of N qudits generated randomly according to the probability distribution $\{\overline{P_i}\}$ is a typical one for a given sequence. One finds that p is given by an expression which is similar to (9), however the binomial factors are replaced by multinomial factors. This is due to the fact that the corresponding distributions are now multinomial instead of binomial. The number of required qubits, $M\log_2(d)$, turns out to be $\log_2(1/p) \sim NI(\{p_k\}, \{\rho_k\})$ (for $N \gg 1$), which leads to the announced compression factor. Also the detailed proof can be carried out in a similar way, replacing the binomial distributions by multionomial distributions and the corresponding Gaussian curves (see e.g. (17)) by multidimensional Gaussians curves.

B. Decoding without knowing the source

Notice that in our protocol for compressing commuting mixed states we have implicitly assumed, in step 2(b) of the decoding stage, that Bob knows which are the eigenvectors of the density matrices, i.e. $|0\rangle$ and $|1\rangle$. This is of course legitimate in any context where both Alice and Bob are provided with a description of the source.

Let us note here that we can slightly modify the protocol in such a way that it works even if Bob does not have such a description. Indeed, suppose that now the eigenstates are $|0'\rangle$ and $|1'\rangle$. All we need is that Alice uses the quantum channel to sent N copies of each of these states. Since the N copies of (say) $|0'\rangle$, $|0'\rangle^{\otimes N}$, are supported on the (N+1-dimensional) symmetric subspace of N qubits, $[log(N + 1)]$ qubits are sufficient to transmit them. For

large N, this does not change the communication cost per qubit, $I(\{p_k\},\{\rho_k\})$. And thus, once Bob has received and decompressed $|\,0'\rangle^{\otimes N}$ and $|\,1'\rangle^{\otimes N}$, he can use single copies of these states to replace the $|\,0\rangle$ and $|\,1\rangle$'s of step 2(b). In this way, he does not need to know the details of the source to prepare faithful sequences $\sigma^B_{\mathcal{K}}$.

C. The Levitin-Holevo bound can not always be reached in a blind protocol

In the previous sections, we showed for commuting density operators that in the visible scenario, the bound for the compression factor given by the Levitin–Holevo function can always be reached. Here, we investigate the *invisible* scenario, i.e. the case where Alice does not know the specific sequence to be sent. We give an example where the Levitin-Holevo bound for the compression factor cannot be reached.

We consider two density operators $\rho_1 = |1\rangle\langle 1|$, $\rho_2 = 1/2\mathbb{1}$ with corresponding probabilities $p_1 = p_2 = 1/2$. We will argue that the achievable compression factor C is given by the entropy of the operator $\rho = \sum_k p_k\rho_k$, $S(\rho) \approx 0.8113$, which should be compared with $I(\{p_k\},\{\rho_k\}) \approx 0.3113$. We will not give a formal proof of this statement, but will rather argue in terms of typical sequences and the corresponding "typical states" (see Sec. II).

If we write the operator $\sigma^A_{\mathcal{K}}$ corresponding to a typical sequence in the basis $\{|i_1\rangle\otimes|i_2\rangle\ldots|i_N\rangle\}$ $(i_j = 0,1)$, the typical states are those with exactly $3N/4$ ones (and $N/4$ zeros). Note that Alice can determine with help of a measurement of all qubits in the computational basis which of the typical states she possesses. This can be done without disturbing the signal because she measures in the eigenbasis of $\sigma^A_{\mathcal{K}}$. Let us thus assume that Alice knows the typical state she has to transmit. We can take without loss of generality the state

$$|a\rangle = \underbrace{|1\rangle\otimes|1\rangle\otimes\ldots\otimes|1\rangle\otimes|1\rangle}_{3N/4}\otimes\underbrace{|0\rangle\otimes\ldots\otimes|0\rangle}_{N/4} \quad (24)$$

However, in contrast to the visible case, Alice does not know to which specific (typical) sequence the state $|a\rangle$ belongs to. In fact, there are many sequences which are compatible with the state $|a\rangle$, namely all those which have all $N/2$ density operators ρ_1 at the positions $1,\ldots,3N/4$.

We will show now that the state $|a\rangle$ has to be transmitted "perfectly" to Bob, since even a small derivation from the state $|a\rangle$ will lead to a macroscopic error. To this aim, we consider a general coding/decoding procedure. Notice that Bob can measure in the computational basis after decoding the received signal and thereby obtain with some probability a pure state $|b\rangle$ [17] which is a sequence of zeros and ones. Let us assume that $|b\rangle$ differs from $|a\rangle$ only at two positions, e.g. the first and the N^{th} qubits are flipped (note that two states must always differ at an even number of positions, as we assumed that the total number of zeros/ones is fixed). The average error can be written as follows

$$E = \sum P(\sigma^A_{\mathcal{K}}/a)E(b,\sigma^A_{\mathcal{K}}), \quad (25)$$

where the sum runs over all possible typical sequences $\sigma^A_{\mathcal{K}}$, $P(\sigma^A_{\mathcal{K}}/a)$ is the probability that we deal with sequence $\sigma^A_{\mathcal{K}}$ provided that Alice possesses the state $|a\rangle$ and $E(b,\sigma^A_{\mathcal{K}})$ is the error for the sequence $\sigma^A_{\mathcal{K}}$ given that Bob received the state $|b\rangle$. Under our previous assumption on $|a\rangle,|b\rangle$, we have that $E(b,\sigma^A_{\mathcal{K}})$ is either one (for all sequences which have ρ_1 at position one) or zero (for all sequence which have ρ_2 a position one). As there are

$$q \equiv \left(\begin{array}{c} 3N/4 \\ N/2 \end{array}\right) \quad (26)$$

sequences which are compatible with $|a\rangle$, we have that $P(\sigma^A_{\mathcal{K}}/a) = 1/q$ for all those sequences and zero otherwise. It is easy to see that

$$E = \left(\begin{array}{c} 3N/4-1 \\ N/2-1 \end{array}\right) / \left(\begin{array}{c} 3N/4 \\ N/2 \end{array}\right) = 2/3, \quad (27)$$

i.e. the average error is already macroscopic even when $|b\rangle$ differs from $|a\rangle$ only at two positions. We conclude that in order to have E sufficiently small (and thus the fidelity sufficiently close to one), we must have that $|b\rangle = |a\rangle$. This implies that *all* typical states have to be transmitted perfectly from Alice to Bob, as our analysis is not restricted to the specific choice of $|a\rangle$. There are

$$g \equiv \left(\begin{array}{c} N \\ 3N/4 \end{array}\right) \quad (28)$$

typical states, which means that $\log_2(g) \sim NS(\rho) \approx 0.8113N$ qubits are required for perfect transmission and no further compression is possible [19]. Thus, the Levitin-Holevo bound cannot be reached in this case. On the other hand, if $p_1 = \epsilon \to 0$, it happens that —also in the invisible scenario— the achievable compression factor approaches $I(\{p_k\},\{\rho_k\}) \to 0$, while $S(\rho) \to 1$.

Note that this analysis is not restricted to this specific example but can be generalized to determine the compression factor C, $S(\rho) \geq C \geq I(\{p_k\},\{\rho_k\})$, also in the invisible case.

V. SUMMARY

We have analyzed the compression of mixed states in the visible case and for commuting density operators. We have given a protocol that achieves the compression factor (4), which was known to be a lower bound. Our protocol is based on the creation of the same set of random numbers by Alice and Bob, and choosing among them the one appropriated to the sequence they want to send. This protocol can be extended to the case in which the density operators do not commute. In that case, Alice and Bob can encode the states in the same random subspaces within the typical subspace. This problem will be addressed in a future work.

ACKNOWLEDGMENTS

We thank C. Fuchs for interesting discussions. This work was supported by the Austrian Science Foundation under the SFB "control and measurement of coherent quantum systems (Project 11), the European Community under the TMR network ERB–FMRX–CT96–0087 and project EQUIP (contract IST-1999-11053), the European Science Foundation, and the Institute for Quantum Information GmbH. G.V also acknowledges funding from the EC through grant No. HPMF-CT-1999-00200.

APPENDIX A: BOB'S DENSITY OPERATOR AND FIDELITY

We denote by $\{|\Psi_m\rangle\}_{m=1}^{2^N}$ the computational basis for the N qubits of the sequence, i.e. $|\Psi_1\rangle = |0,0,\ldots,0\rangle$, $\ldots,|\Psi_{2^N}\rangle = |1,1,\ldots,1\rangle$. According to the protocol given in Section III, Bob's density operator can be written as follows:

$$\sigma_{\mathcal{K}}^B = \sum_{x_{1,2}=0}^{n_{1,2}} P(x_{1,2}) \sum_{t=1}^{S} P(t, x_{1,2}, v_{1,2}) \times \sum_{m=1}^{2^N} P(m/t, x_{1,2}, v_{1,2}) |\Psi_m\rangle\langle\Psi_m| + \sum_{x_{1,2}=0}^{n_{1,2}} P(x_{1,2}) E(x_{1,2}, v_{1,2}) \frac{\mathbb{1}}{2^N}. \quad \text{(A1)}$$

Here, $P(x_{1,2})$ is the probability that Alice obtains x_1 and x_2 and is given in (12); $P(t, x_{1,2}, v_{1,2})$ is the probability that among the S random sequences, there are t with exactly $x_{1,2}$ ones (and the rest zeros) at the positions indicated by $v_{1,2}$; $E(x_{1,2}, v_{1,2}) \equiv P(0, x_{1,2}, v_{1,2})$, i.e. the probability that the error state is produced; $P(m/t, x_{1,2}, v_{1,2})$ is the probability that given t sequences with exactly $x_{1,2}$ ones (and the rest zeros) at the positions indicated by $v_{1,2}$, and we choose one of them randomly, Bob obtains the sequence of zeros and ones corresponding to $|\Psi_m\rangle$. This last can be reexpressed as

$$P(m/t, x_{1,2}, v_{1,2}) = \sum_{x=0}^{t} \binom{t}{x} P(m/x_{1,2}, v_{1,2})^x \times [1 - P(m/x_{1,2}, v_{1,2})]^{t-x} \frac{x}{t} = P(m/x_{1,2}, v_{1,2}), \quad \text{(A2)}$$

where $P(m/x_{1,2}, v_{1,2}) \equiv P(m/1, x_{1,2}, v_{1,2})$. Now, we can perform the sum over t in (A1) and obtain

$$\sigma_{\mathcal{K}}^B = \sum_{x_{1,2}=0}^{n_{1,2}} P(x_{1,2})[1 - E(x_{1,2}, v_{1,2})] \times \sum_{m=1}^{2^N} P(m/x_{1,2}, v_{1,2}) |\Psi_m\rangle\langle\Psi_m| + \sum_{x_{1,2}=0}^{n_{1,2}} P(x_{1,2}) E(x_{1,2}, v_{1,2}) \frac{\mathbb{1}}{2^N}. \quad \text{(A3)}$$

On the other hand, we can write

$$\sigma_{\mathcal{K}}^A = \sum_{x_{1,2}=0}^{n_{1,2}} P(x_{1,2}) \sum_{m=1}^{2^N} P(m/x_{1,2}, v_{1,2}) |\Psi_m\rangle\langle\Psi_m|. \quad \text{(A4)}$$

The fidelity $F(\sigma_{\mathcal{K}}^A, \sigma_{\mathcal{K}}^B)$ will be larger than or equal to the one calculated by ignoring the term proportional to the identity operator in (A3). We obtain

$$F(\sigma_{\mathcal{K}}^A, \sigma_{\mathcal{K}}^B) \geq \sum_{x_{1,2}=0}^{n_{1,2}} \sum_{m=1}^{2^N} P(x_{1,2}) P(m/x_{1,2}, v_{1,2}) \times [1 - E(x_{1,2}, v_{1,2})]^{1/2} \geq 1 - E_{\mathcal{K}}, \quad \text{(A5)}$$

where

$$E_{\mathcal{K}} = \sum_{x_{1,2}=0}^{n_{1,2}} P(x_{1,2}) E(x_{1,2}, v_{1,2}), \quad \text{(A6)}$$

and we have used

$$\sum_{m=1}^{2^N} P(m/x_{1,2}, v_{1,2}) = 1. \quad \text{(A7)}$$

Thus, the condition

$$E = \sum_{\mathcal{K}} P_{\mathcal{K}} E_{\mathcal{K}} < \epsilon, \quad \text{(A8)}$$

automatically implies that $\overline{F} > 1 - \epsilon$. Now, both $P_{\mathcal{K}}$ and $E_{\mathcal{K}}$ only depend on the number of times that ρ_1 appears in $\mathcal{K}$, and not on how they are placed, so that we can write (14).

[1] E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[2] B. Schumacher, Phys. Rev. A **51**, 2738 (1995);
[3] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[4] H. Barnum, C.A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996); R. Jozsa, P. Horodecki, M. Horodecki, and R. Horodecki, Phys. Rev. Lett **81**, 1714 (1998).
[5] M. Horodecki, Phys. Rev. A **57**, 3364 (1998).
[6] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, quant–ph/0008024.
[7] M. Horodecki, Phys. Rev. A **61**, 52309 (2000).
[8] In the following we will freely interchange the words sequence and state whenever there is a one to one correspondence.

[9] H.-K. Lo, Optics Comm. **19**, 552 (1995).
[10] P. Hauslanden, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wooters, Phys. Rev. A **54**, 1869 (1996).
[11] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).
[12] B. Schumacher and M. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[13] G. Kramer and S. A. Savari, "Quantum Data Compression of Ensembles of Mixed States with Commuting Density Operators", in preparation.
[14] Note that, according to Ref. [6] we are using the global fidelity as opposed to the local fidelity. However, in the case of commuting operators the result does not depend on the fidelity (both for the blind and visible cases). This can be easily understood as follows. If there exist a protocol which achieves a smaller compression factor using the local fidelity, then Alice and Bob can use the same protocol but in which Bob measures the decoded sequence of qubits in the $\{|0\rangle, |1\rangle\}$ basis. It is clear that this method will not change the local fidelity, but will give a global fidelity equal to the local one. On the other hand, as pointed out in Ref. [6] the local fidelity criterion is less stringent than the global one.
[15] The values $n_{1,2}$ and $v_{1,2}$ depend on the sequence $\mathcal{K}$. However, in order to make the notation simpler we will omit this dependence throughout.
[16] Note that the specific values of K and α are irrelevant for our discussion. Nevertheless, using the extreme values of n_i and x_i given in the intervals (18) and (19), and bounding the resulting expressions, one can easily find, for example, that (20) holds with $K = (8\pi\overline{P}_1\overline{P}_0)^{1/2}$ and $\alpha = 72/(\overline{P}_1\overline{P}_0)$.
[17] It follows from the increasing character of the fidelity under trace preserving operations [18], that if we measure non–selectively the decoded state $\sigma^B_{\mathcal{K}}$ in the computational basis, its fidelity with respect to the initial state $\sigma^A_{\mathcal{K}}$ will not decrease (notice that the initial state $\sigma^A_{\mathcal{K}}$ is left invariant under the measurement). If we now in addition make the same measurement with post selection, we obtain a certain pure state $|b\rangle$, which can be written as a sequence of zeros and ones.
[18] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge University press (2000)
[19] Note that when using a different definition of the fidelity, further compression may be possible. Consider for example the average local fidelity, which is defined as $\overline{F} = 1/N \sum_{l=1}^{N} F(\rho_l, \hat{\rho}_l)$, where ρ_l $[\hat{\rho}_l]$ is the reduced density operator at position l of the original [received] system (sequence). This definition of fidelity is different from both, global and local fidelity as used in [6] and throughout this paper. In that case, the error $E(b, \sigma^A_{\mathcal{K}})$ scales like $2/N$, as errors on two positions only affect two out of N density operators. This suggests that one may allow for certain imperfection in the transmission of state $|a\rangle$ without producing a macroscopic error. Thus not all typical states have to be sent perfectly but one may rather achieve further compression.